



TITLE:

クラッシュコース(数学者のための 分子生物学入門,研究会報告)

AUTHOR(S):

加藤, 毅

CITATION:

加藤, 毅. クラッシュコース(数学者のための分子生物学入門,研究会報告). 物性研究 2003, 81(1): 34-52

ISSUE DATE:

2003-10-20

URL:

<http://hdl.handle.net/2433/97616>

RIGHT:

クラッシュ コース

加藤 毅

1. はじめに

このノートは、`数学者のための分子生物学入門'の初日に数学者向けに行われた講義の内容である。これは次の日から続いた生物学の専門家の講義を理解するために、少なくとも知っているべき基礎知識の一部を解説したものである。内容は中心教義と突然変異に絞り、できるだけ短くした。素人が初学者向けに行った話しであり、中身に多くの不備があることをご了承願いたい。内容はかなり大雑把であり、思いきって省略してあることがたくさんある。

2. 言語

1.A 言語：形式的な言語とは以下で定まる集合 L のことである。まず、アルファベットと呼ばれる有限集合 $A = \{a_1, \dots, a_N\}$ があり、

$$A^* = \{a_{i_1} \dots a_{i_l}; i_1, \dots, i_l \in \{1, \dots, N\}\}$$

をすべてのワードの集合とする。この時、 A から定まる言語とは A^* の部分集合 L のことである。

ここでの主役は、次の三つの言語である。

- (1) DNA (四つのアルファベット, A, T, G, C から成る)
- (2) RNA (四つのアルファベット, A, U, G, C から成る)
- (3) タンパク質 (二十個のアルファベットから成る)

ここで二十個のアルファベットの各元はアミノ酸と呼ばれる。

1.B 中心教義：分子生物学の中心教義とよばれるものが、次のものである。

- (1) : 言語の間に二つのユニバーサルな写像、

転写: $\text{DNA} \rightarrow \text{RNA}$, 翻訳: $\text{RNA} \rightarrow \text{タンパク質}$

が存在する。

- (2) : 自己複製と呼ばれる、同じ言語を二つ作る機構が存在する。

$\text{DNA} \rightarrow \text{DNA} \cup \text{DNA}.$

タンパク質は、機能を持つ、なにか実体の伴ったものである。一方 DNA は単なる記号列と見れば、

$$\rho: \text{DNA} \mapsto \text{タンパク質}$$

は自然の作り上げた一つの表現と言える。

$a \in \text{DNA}$ から $\rho(a)$ へ至る過程で、化学反応をおこすことで様々な中間段階を経る。ここで、酵素と呼ばれるタンパク質がその化学反応を促進または抑制する。触媒機能を持つと言う。従って、図式、

$$a - [x_1] \rightarrow a_1 - [x_2] \rightarrow a_2 - [x_3] \rightarrow \dots a_{n-1} - [x_{n-1}] \rightarrow a_n = \rho(a)$$

のように現される。ここで各 x_i は酵素、よってタンパク質である。一方、各 x_i ももとをただせばある DNA の元 y_i からできている ($x_i = \rho(y_i)$) から、上の図式よりは、次のもののほうがよい。

$$\{a, y_1, \dots, y_{n-1}\} - [\rho(y_1)] \rightarrow a_1 - [\rho(y_2)] \rightarrow a_2 \dots$$

問：一連の相互作用をアブストラクトに記述する。

この問いに向かうためにはどのような数学の道具が使えるであろうか。まず、オートマトンを考えることは自然であろう。

1.B 有限状態オートマトン： より一般に考えるならばチューリングマシンを扱うべきであるが、数学、特に幾何学に関係するのが有限状態オートマトンであるので、ここではそれに限った。

有限状態オートマトンとは、以下のデータから構成される。

$$A = \{a_1, \dots, a_n\} : \text{アルファベット}$$

$$S = \{s_0, \dots, s_l\} : \text{状態集合, } Y \subset S : \text{受け入れられる状態}$$

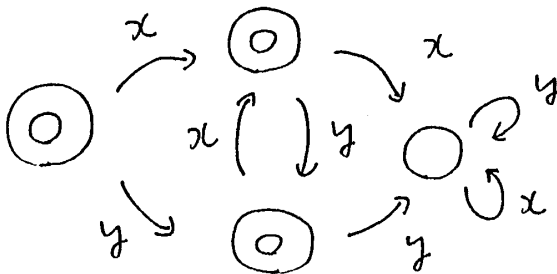
$$\mu: S \times A \mapsto S.$$

以上のデータから A 上の言語 L が以下で得られる。

$$L = \{a_{i_1} \dots a_{i_m} \in A^* : \mu(s_0, a_{i_1}) = s_{i_1}, \dots$$

$$\mu(s_{i_j}, a_{i_{j+1}}) = s_{i_{j+1}}, \dots, \mu(s_{i_{m-1}}, a_{i_m}) \in Y\}.$$

例：次のオートマトンは \mathbb{Z}_2 の自由積、 $\mathbb{Z}_2 * \mathbb{Z}_2$ を現している。



$$\begin{aligned} \odot &\in Y \\ x &\in \mathbb{Z}_2, y \in \mathbb{Z}_2 \\ L &= \mathbb{Z}_2 * \mathbb{Z}_2 \end{aligned}$$

例：有限生成群におけるオートマトン問題： Γ を有限生成群、 $A = \{a_1, \dots, a_n\}$ を生成元とする。任意のワード $w = a_{i_1} \dots a_{i_m}$ に対して Γ の元 $\pi(w) \in \Gamma$ が自然に与えられる。

問：オートマトン $M = (A, S, Y, \mu)$ であって、

- (1) $\pi: L(M) \mapsto \Gamma$ が全射、 (2) ‘良い性質を持つ’

ものが存在するか？

(2) の例としてワードが定める道に幾何学的な条件を付けたものが良く知られている。この場合そのような条件を満たすオートマトンを持つ群をオートマチック群と呼ぶ。オートマチック群であるためには強い条件を満たさねば成らない（語の問題が解ける、等周不等式のオーダーが2以下、等）（[ECHLPT]）。

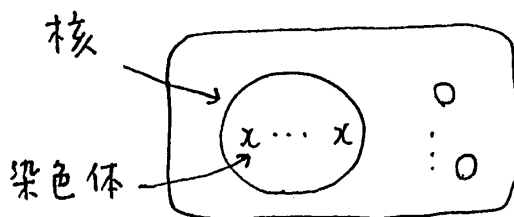


一方で、酵素は相互作用することでシステムが成り立っているが、相互作用を含めたオートマトンの解析がここでは必要になろう。さらに、先走ってしまえば突然変異を含めたものも考える必要がある。

3. DNA

生物は二種類、原核生物と真核生物に分けられる。ヒトは真核生物に属する。真核生物の細胞には核があり、その中に染色体が含まれている。染色体には遺伝情報が含まれている。

真核生物の細胞



核の直径

$$\approx 10 \mu m = \frac{1}{100万} m$$

染色体は、DNA とタンパク質から成り、大雑把に言ってタンパク質はヒストンと呼ばれるものである。

DNA と RNA を構成するアルファベットに対応する分子は、それぞれ次のように名前が付いている。

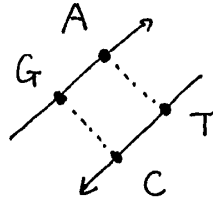
DNA: A: アデニン、T: チミン、G: グアニン、C: シトシン

RNA: A: アデニン、U: ウラシル、G: グアニン、C: シトシン

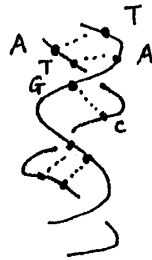
DNA についてはまず、次のことを知っていなければならない。

FACTS: (1) : DNA は非常に長い二本のストリング

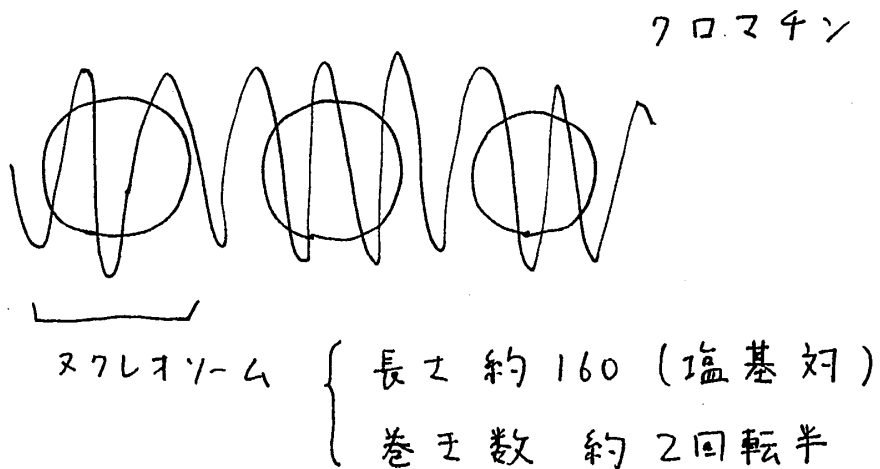
(2) : DNA は互いに逆向きに向きつけられていて、向き合う相手は決まっている。



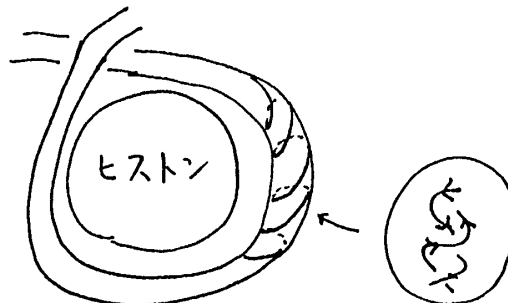
(3) : DNA はらせん構造をもつ。



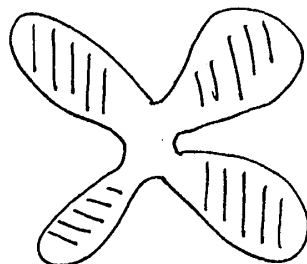
(4) : DNA はヒストンに巻き付いている。



(5) : さらに DNA は超らせんを形成。



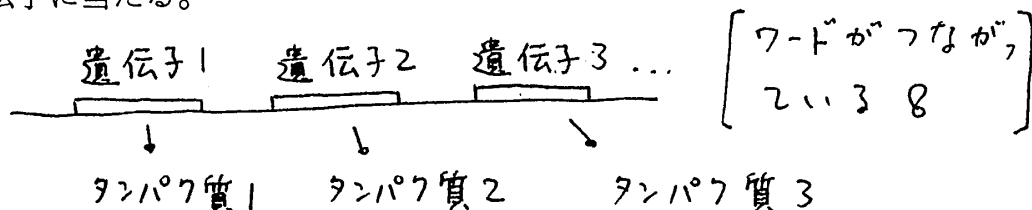
染色体はクロマチンが非常に小さく折畳まったもの。



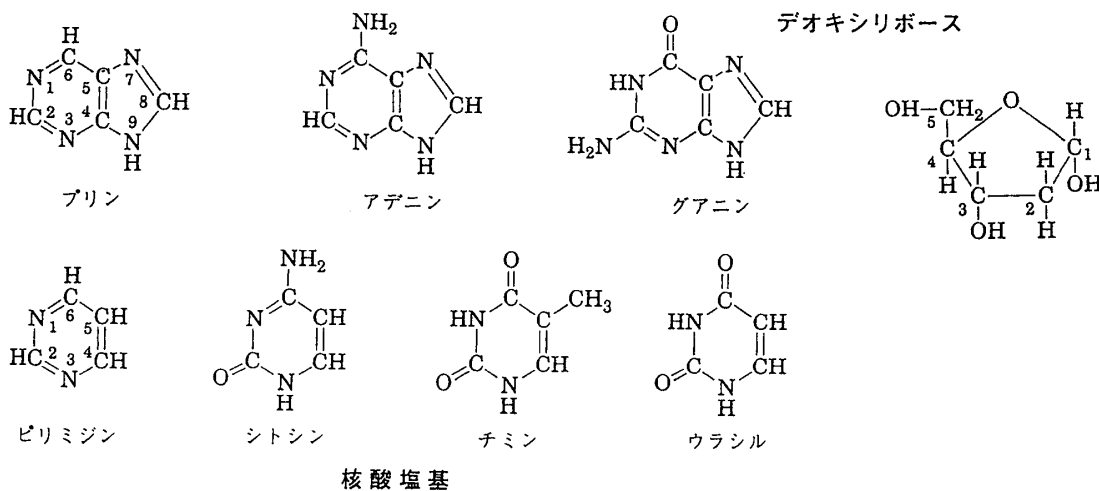
約 $\frac{1}{8400}$ の長さ

問： クロマチンの凝縮と分散のメカニズム？

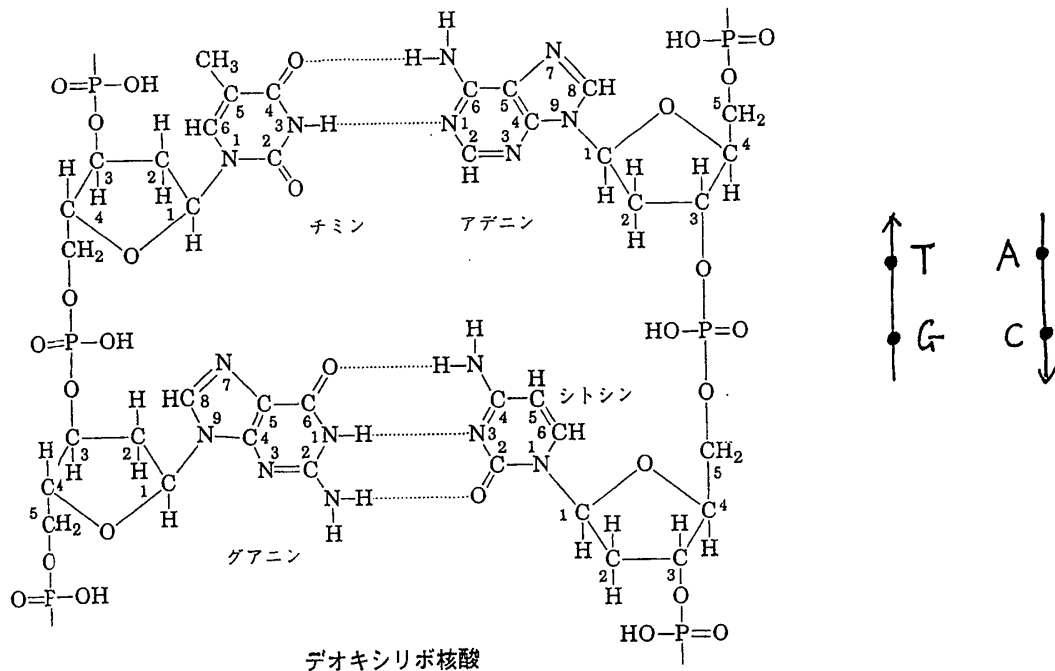
注： (1) DNA は言語であって長いストリングである。大雑把に言って、各ワードが遺伝子に当たる。



(2) もう少し詳しく分子の構造を書くと、DNA や RNA の構成分子は、塩基と呼ばれる、アルファベットによって異なる分子、(デオキシ)リボースと呼ばれる糖とリン酸である。各アルファベットにはそれぞれ三つの分子、(塩基、(デオキシ)リボース、リン酸) が対応する。



DNA の向きについて。デオキシリボースの炭素には番号が付いていて、向きは、5 番目の炭素から 3 番目に向かうように入れる。5' → 3' と書く。



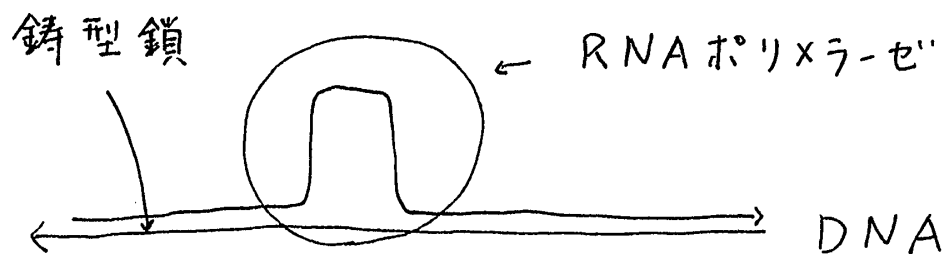
4. 中心教義 (転写)

転写とは、言語間の写像、

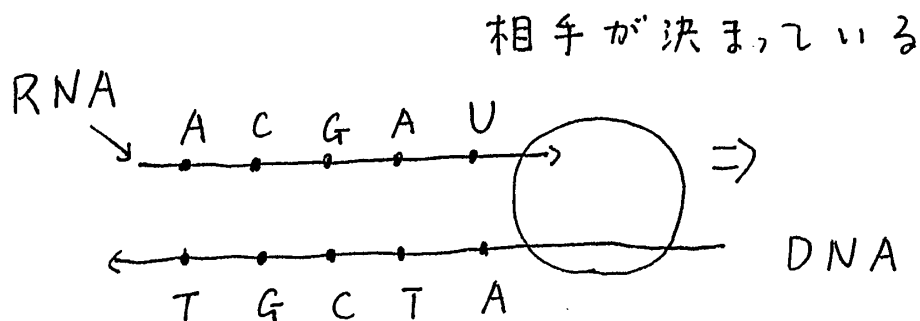
DNA ⇌ RNA

のことであった。DNA のワードから RNA のワードを作る働きをするものが、RNA ポリメラーゼと呼ばれる酵素である。その仕組みは大雑把に言って以下のようである。

① RNA ポリメラーゼが結合



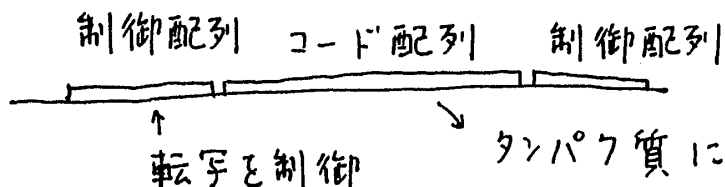
② 転写開始



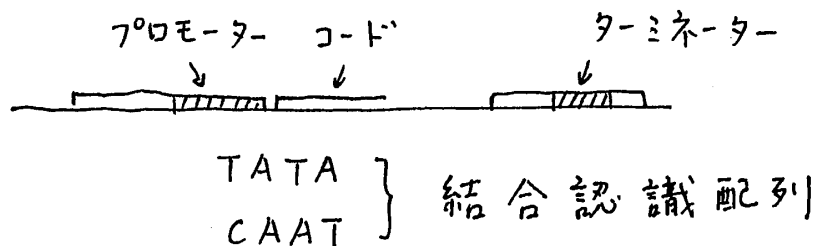
3.B 制御配列： 実は DNA のすべてのワードが転写されるわけではない。タンパク質になる遺伝子=ワードと同じくらい大事な役目を担うが転写されないワードのなかに、制御配列と呼ばれるワードがある。よって、DNA のワードは制御配列とコード配列に分けられる。

コード配列： RNA やタンパク質になる部分、

制御配列： いつ、どのくらいの量を転写するか調節する部分。



制御配列の中にプロモーターとターミネーターと呼ばれる配列があり、転写の際 RNA ポリメラーゼはまずプロモーターに結合し、ターミネーターで終了する。



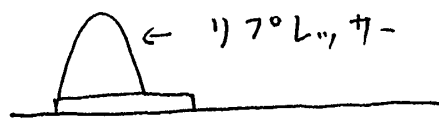
プロモーターの中で、真核生物では TATA、大腸菌では CAAT (プリブナウ配列) の配列が結合の際の目印になっている。

転写を調節する酵素は転写調節因子と呼ばれ、制御配列に結合する。

リプレッサー： 転写を妨げる (オペレーターに結合)

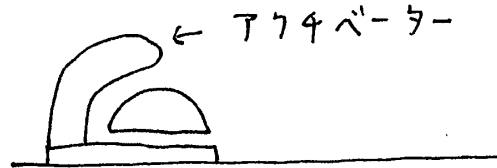
アクチベーター： 転写を促進 (プロモーターに結合)

(a)



(RNA ポリメラーゼ が)
(結合 できない)

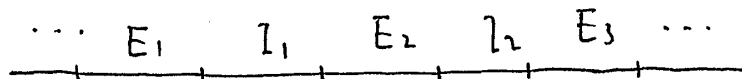
(b)



(RNA ポリメラーゼ)
(が 結合 し や す い)

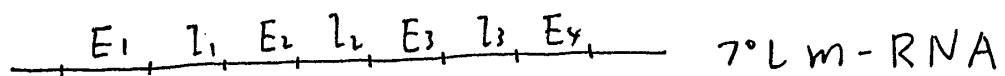
転写されてできたものは、メッセンジャー RNA (m-RNA) と呼ばれるが、真核生物の場合、正確にはこれはプレ m-RNA と呼ばれる。もう一段階のステップを経て m-RNA が出来上がるのである。そのステップはスプライシングと呼ばれる。

7°L m-RNA



5. スプライシング

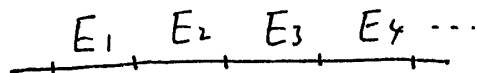
真核生物の場合、プレ m-RNA は二つの部分に分けられる。



7°L m-RNA

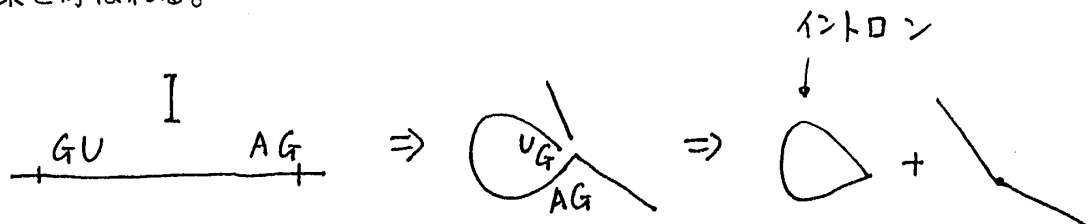


スプライシング



m-RNA

各 E_i はエクソンと呼ばれ、タンパク質に翻訳される部分で、各 I_i はイントロンと呼ばれ切り捨てられる部分である。切り捨てる働きをする酵素はスプライシング酵素と呼ばれる。



問： イントロンの意味？

6. 翻訳

翻訳とは、言語間の写像、

RNA \mapsto タンパク質

のことであった。タンパク質は二十個のアミノ酸からなる配列である。

{A, U, G, C} の各三文字に対してアミノ酸一つが対応する。これをコドン対応と言う。ここで大事なことは、この対応が基本的にはすべての生物に共通していることである。例えば、 β グロビン遺伝子は、

GTT CAC CTC ACT CCC GAG GAA AAG
Val His Leu Thr Pro Glu Glu Lys (6.1)

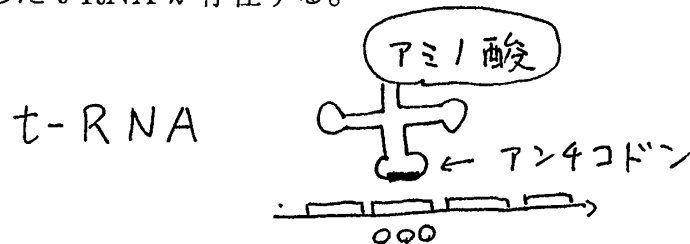
コドン対応は一對一ではなく、同じアミノ酸を現す三つ組みを同義コドンと呼ぶ。以下に普遍的な対応を現す、遺伝暗号表を載せる。

第一塩基	第二塩基				第三塩基	アミノ酸			
	U	C	A	G		アミノ酸	一文字略号	三文字略号	mRNA のコード
U	Phe	Ser	Tyr	Cys	U	アラニン	A	Ala	GCU, GCC, GCA, GCG
	Phe	Ser	Tyr	Cys	C	アルギニン	R	Arg	CGU, CGC, CGA, CCG, AGA, AGG
	Leu	Ser	終止	終止	A	アスパラギン	N	Asn	AAU, AAC
	Leu	Ser	終止	Trp	G	アスパラギン酸	D	Asp	GAU, GAC
C	Leu	Pro	His	Arg	U	システイン	C	Cys	UGU, UGC
	Leu	Pro	His	Arg	C	グルタミン酸	E	Glu	GAA, GAG
	Leu	Pro	Gln	Arg	A	グルタミン	Q	Gln	CAA, CAG
	Leu	Pro	Gln	Arg	G	グリシン	G	Gly	GGU, GGC, GGA, GGG
A	Ile	Thr	Asn	Ser	U	ヒスチジン	H	His	CAU, CAC
	Ile	Thr	Asn	Ser	C	イソロイシン	I	Ile	AUU, AUC, AUA
	Ile	Thr	Lys	Arg	A	ロイシン	L	Leu	UUA, UUG, CUU, CUC, CUA, CUG
	Met, fMet	Thr	Lys	Arg	G	リシン	K	Lys	AAA, AAG
G	Val	Ala	Asp	Gly	U	メチオニン	M	Met	AUG
	Val	Ala	Asp	Gly	C	フェニルアラニン	F	Phe	UUU, UUC
	Val	Ala	Glu	Gly	A	プロリン	P	Pro	CCU, CCC, CCA, CCG
	Val	Ala	Glu	Gly	G	セリン	S	Ser	UCU, UCC, UCA, UCG, AGU, AGC
						トレオニン	T	Thr	ACU, ACC, ACA, ACG
						トリプトファン	W	Trp	UGG
						チロシン	Y	Tyr	UAU, UAC
						バリン	V	Val	GUU, GUC, GUA, GUG

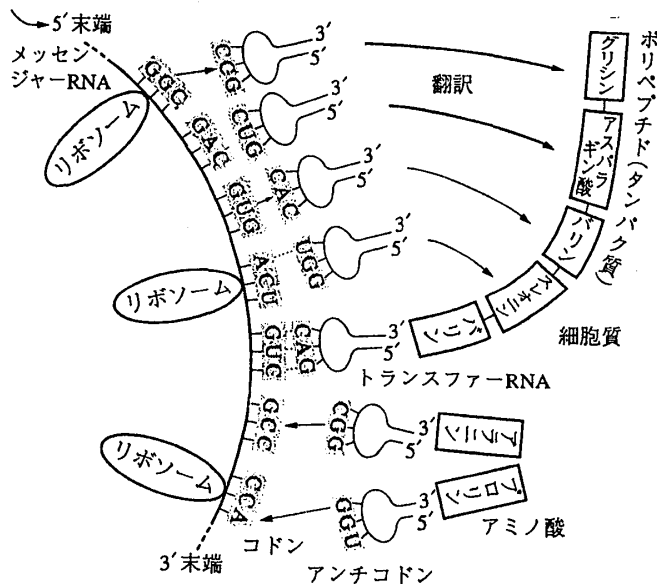
[5] 57

翻訳は転写とは違ったプロセスを経て行われる。m-RNAはタンパク質になるものであった。m-RNAの他に、タンパク質にはならないRNAが二種類存在し、t-RNA（転移RNA）、r-RNA（リボソームRNA）と呼ばれる。よってRNA言語は三種類のワード（mRNA, t-RNA, r-RNA）に分けられる。

翻訳はアダプターを仲介する。その役目はt-RNAが担う。各アミノ酸にそれぞれ対応したt-RNAが存在する。



m-RNAをタンパク質に翻訳する際、どこから始めるか、その目印が決まっています。配列はAUGである。開始コドンと呼ばれる。



AUG: 開始コドン

核のなかでつくられたメッセンジャーRNAが細胞質の中に出てくると、その3つの塩基の組（コドン）に応じて、それと相補的な塩基の3つの組（アンチコドン）を一方にもち、他方にアミノ酸をもっているトランスファーRNAが結合する。これらのアミノ酸が重合してタンパク質ができる

[宮]より

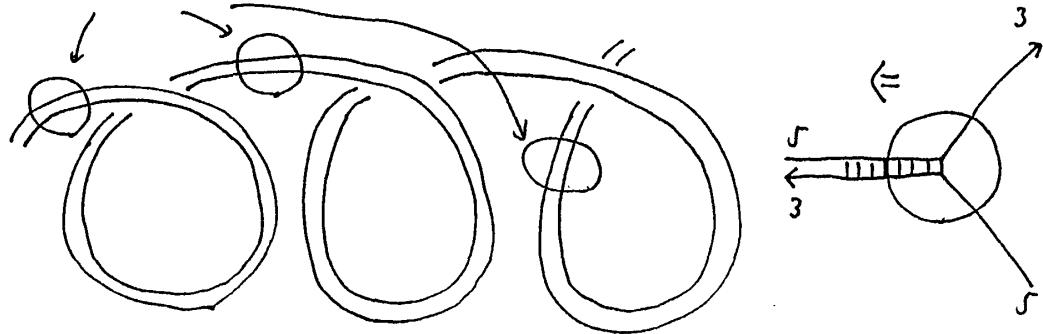
7. 自己複製

自己複製は、DNAと同じものを二つ作るプロセスのことである。

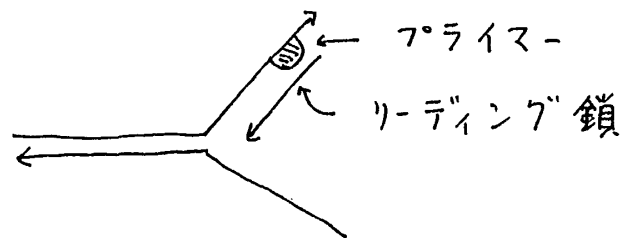
DNA \mapsto DNA \cup DNA.

ここでは、大雑把に言うことにし、ヘリカーゼ、DNA ポリメラーゼと呼ばれる二つの酵素と、プライマーと呼ばれる RNA が複製の働きをする。以下プロセスの概略を述べる。

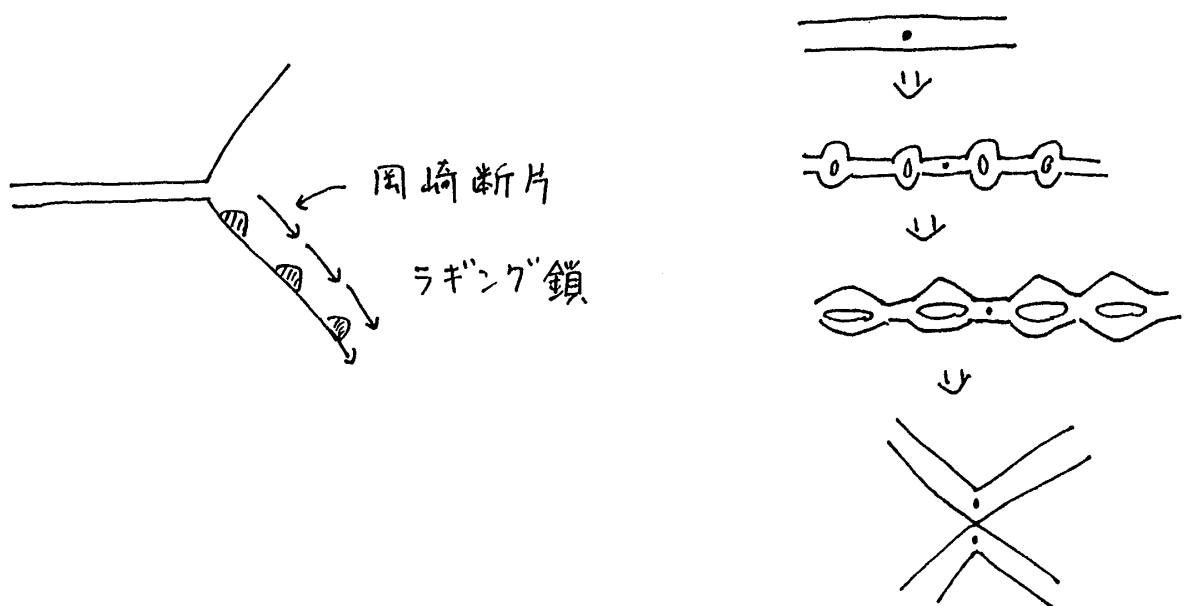
(1) ヘリカーゼがらせんを順にほぐす。



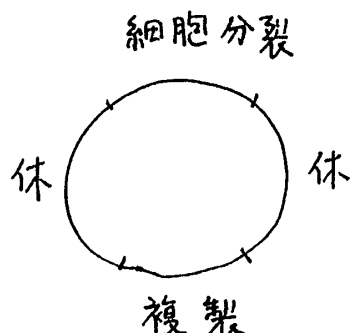
(2) 複製の始点を決めるプライマーが結合し、DNA ポリメラーゼが相補的に複製していく。



(3) ほぐしながら複製するが、ほぐす向きと複製する向きが同じ時、それをリーディング鎖、逆向きをラギング鎖と呼ぶ。ラギング鎖のほうはいかにも作りにくそうだが（後ろ向きに歩いているように）、この場合、プライマーが少しずつ前のほうに付いていくことで鎖を延ばしていく。その断片は岡崎断片と呼ばれる。



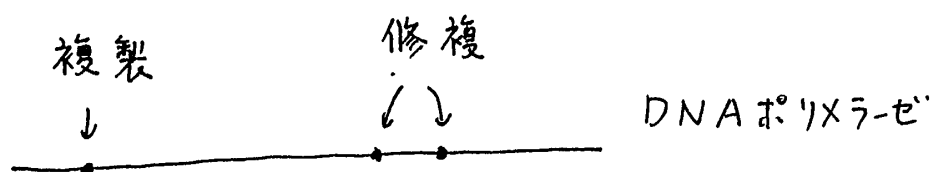
このようにして出来上がった二つの DNA は細胞が分裂することで分けられる。ここで真核生物の場合、細胞分裂の時期と複製の時期がくっきり別れる。複製が終わると、しばらく休んでから分裂し、またしばらく休んでから複製を始める。非連続的なプロセスである。これを細胞周期と呼ぶ。



このようにして細胞は増えていくが、複製していく段階で間違った配列を作ったりしないのだろうか。実はほぼ完全な複製が行われている（せいぜい 100 万 - 1000 万塩基に一個のエラー）。そんなに完全な機構が存在するのだろうか。実は、DNA ポリメラーゼはよく間違えるが、それらを直してしまう、二種類の修復機構を持っている。

FACT: DNA ポリメラーゼは、複製と二種類の修復機構をあわせ持つ。

なぜ、このような複数の相異なる機能を持ちうるかというと、DNA ポリメラーゼは巨大な分子で、三ヶ所の違った場所で、違った形を持つ。機能は、形（と極性）によって違いをもたらす。



8. 言語とシフト

言語は有限の列から成り、数学的な解析が極めて難しい。そこでもう少し扱いやすいもので近似することを考えてみる。一つの方法は、有限の列を無限の列で置き換えることである。そのような列の集合をシフトと呼ぶ。

$A = \{a_1, \dots, a_n\}$ をアルファベットとする。

$$\Sigma = \Sigma(A) = \{(a_0, a_1, \dots) : a_i \in A\}$$

を（片側）フルシフトと呼ぶ。ここには自然な位相が入り、コンパクトで距離を入れることができる。

$M \in \text{Mat}_n(\{0,1\})$ を $\{0,1\}$ 行列とすると、 $A = \{1,2,\dots,n\}$ をアルファベットとする、無限列言語が以下で作れる。

$$\Sigma_M = \{(a_0, a_1, \dots) : a_i \in A, M_{a_i, a_{i+1}} = 1\}$$

有限型のシフトと呼ぶ。

例： $f : I = [0, 1] \mapsto [0, 1]$ を次の PL 写像とする。区間は $I = [0, \frac{1}{2}] \cup [\frac{1}{2}, 1] \equiv I_0 \cup I_1$ と分けられ、 $p : I \mapsto \{0,1\}$ が $\frac{1}{2}$ 以外の点で定義される。

今、 $\sigma : I \mapsto \Sigma(\{0,1\})$ を

$$\sigma(x) = (p(x), \dots, p(f^i(x)), p(f^{i+1}(x)), \dots) \in \Sigma(\{0,1\})$$

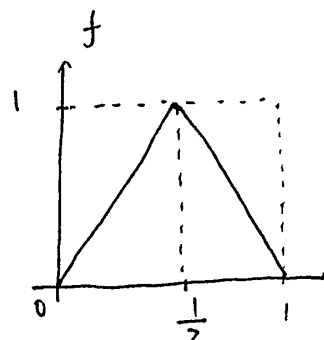
で与える。 σ はすべての点で定義はされないが、定義域は I の中で稠密にとれる。さらに、その逆写像は well-defined に定義され、

$$\pi = \sigma^{-1} : \Sigma(\{0,1\}) \mapsto I$$

が写像として定義される。次の性質を持つ。

- (1) 連続で全射、
- (2) ほとんどの点で 1 対 1、
- (3) 一般に一点の逆像は有限集合。

これは、次のことを示唆している。



言語は微分できる空間で扱える場合がある。

一般の多様体 M とその上の微分同相写像 $f : M \cong M$ に対しても、似たことが言える。今、 $P \subset M$ を周期点の閉包とする。もし f が '良い' 性質を持っているなら (公理 A)、 $P = P_1 \cup \dots \cup P_l$ と分けられ、各 P_i はある有限型シフトに対して上と同様な性質を持つ。これにはマルコフ分割を用いる ([K] 参照)。

9. 酵素

これまでにいくつかの酵素があらわれた。ポリメラーゼ、転写調節因子、、、。これらが一体どのようにしてしかるべき場所、活性化ドメイン、に結合できるのだろう。キイは、

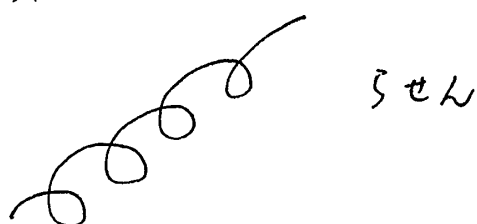
三次元構造：パネルの表と裏のようにうまく形が重なる、

ことである。形から結合する場所が特異的になる。このことから、形を知ることが極めて重要であることが分かる。立体構造は、ある程度のパターンがあり、少なくとも 3 次元構造まで、分類されている。

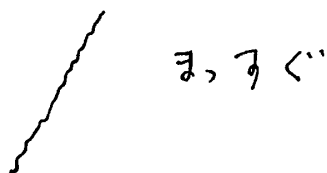
- (1) 一次構造＝アルファベット列： $a_1 \dots a_l$ 。
- (2) 二次構造＝局所的な形：局所的にタンパク質は、おおまかに言って 4

種類の構造に分けられる。

α ヘリックス



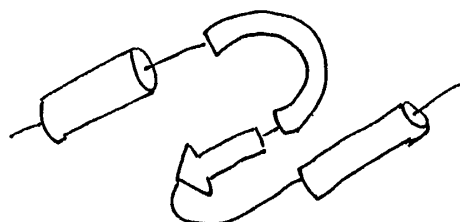
β シート



ターン、ループ



(3) 三次構造=二次構造の組み合わせ：



ドメインを形成

アンフィンゼンの一意性仮説は、タンパク質の立体構造は、配列だけから決まる、ことを述べている。

予測問題：(1) タンパク質の配列だけから立体構造を予測できるか？

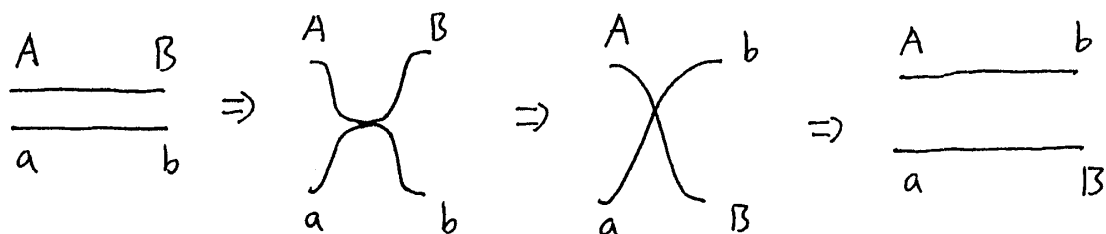
(2) 立体構造から機能を予測できるか？

どちらの問題についても、アイデアは次のようである。 A と B のタンパク質配列を知っていて、それらが互いに似ているとする。さらに A の立体構造（機

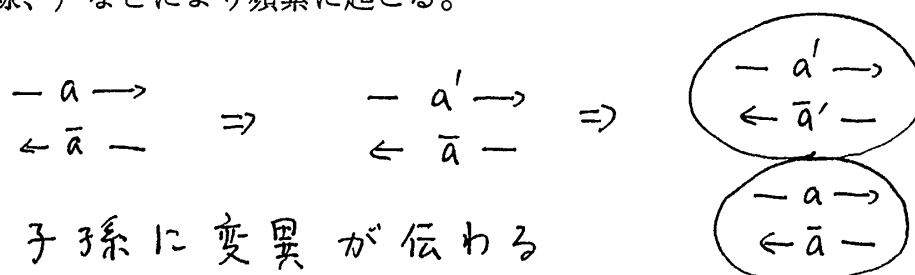
能) が良く分かっているならば、 B も A と似た立体構造 (機能) をもつであろう。似た配列を探す作業をホモロジー検索と呼ぶ。

10. 突然変異

DNA 配列の変化には二種類あり、組み換え、突然変異とそれぞれ呼ばれる。組み換えについて、例えば真核生物では生殖細胞の減数分裂時などに見られる。



突然変異 (mutation) は、塩基配列に変化を与えるもので、電離放射線 (紫外線、X線、) などにより頻繁に起こる。



突然変異は分類されているので、それを以下に書いてみる。

(1) 点突然変異:

$$a_1 \dots a_l \rightarrow a_1 \dots a_{m-1} a'_m a_{m+1} \dots a_l.$$

(2) 欠失変異:

$$a_1 \dots a_l \rightarrow a_1 \dots a_m a_{m+k} \dots a_l.$$

(3) 挿入変異 (特に重複):

$$a_1 \dots a_l \rightarrow a_1 \dots a_m b_1 \dots b_k a_{m+1} \dots a_l.$$

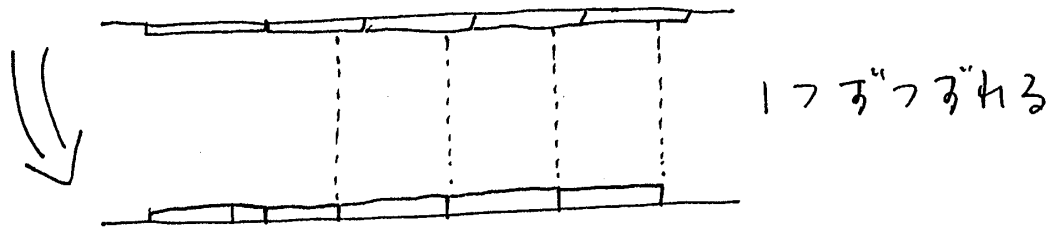
(4) 逆位:

$$a_1 \dots a_l \rightarrow a_1 \dots a_m b_{m+1} \dots b_{m+k} a_{m+k+1} \dots a_l$$

ここで $b_{m+1} \dots b_{m+k} = a_{m+k} \dots a_{m+1}$ 。

突然変異によって、起こる深刻な影響の一つに、フレームシフトと呼ばれるものがある。

コドン対応は、三つで一組だったので、どこから読みはじめるか、によって出来上がるアミノ酸配列が全く異なる。



例 1 :

野生型 :	GTT	CAC	CTC	ACT	CCC	GAG	GAA	AAG
	Val	His	Leu	Thr	Pro	Glu	Glu	Lys
置換 :	GTT	<u>AAC</u>	CTC	ACT	CCC	GAG	GAA	AAG
		Asn						
挿入 :	GTT	CAC	CTC	ACT	CCC	<u>GAC</u>	GGA	AAA
						Asn	Gly	Lys ^(10.1)
逆位 :	GTT	<u>CCC</u>	<u>ATC</u>	ACT	CCC	GAG	GAA	AAG
		Pro	Ile					

例 2 :

野生型 :	CAG	TTT	GAA	CGC	AAA
	Gln	Phe	Glu	Arg	Lys
ナンセンス (置換) :	CAG	TTT	<u>TAA</u>	CGC	AAA
			停止		
ミスセンス (置換) :	CAG	<u>TTA</u>	GAA	CGC	AAA
		Leu			
挿入 :	CAG	TTT	<u>GA</u>	ACG	CAA
			AAG G	Gly	Thr
欠失 :	CAG	<u>TTT</u>	<u>GAAC</u>	GCA	AAC
			Asn	Ala	Asn

配列の中で、突然変異の起こりやすいものとそうでないものがある。大事な機能を担うタンパク質に対応する遺伝子は突然変異しにくい。これを機能的制約と呼ぶ。一方で突然変異を起こしてもあまり機能に支障がない例をあげる。

偽遺伝子：機能を持たない遺伝子（遺伝子が重複した時、それが持つ機能はどちらか一方が担えば良いので、もう一方は変異しても差し支えない、ので突然変異をくり返し、ついには機能を失ったもの）。

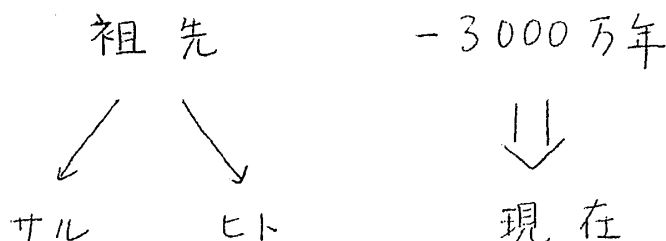
同義突然変異： コドン対応は1対1でなかった。UUUとUUCは同じフェニルアラニンを現している。配列の中で、UUUがUUCに変わっても同じアミノ酸を作るので、その変異は機能に影響を与えない。同義コドンは第三番目の座位のみが違う場合が多い。

以下、点突然変異のみ考える（[宮]参照）。ヒトとさるが持つインシュリン遺伝子は互いに少しだけ配列に差がある。どちらの種も祖先の配列とは違った配列を持つ部分は多重置換と呼ばれる。

祖先	TGCGGC	T	CGCACCT	G	GTGTGGA	
ヒト		A		A		(10.3)
サル		C		G		

多重置換 サルだけ置換

ヒトとさるは約3000万年前に分岐したことが分かっていた。それらの祖先を考えると次の図を得る。



多重置換は考えないことにすると、変異数は2で、配列の長さは20。よって変異率は、

$$\text{変異率（相違度）} = \frac{2}{20} = 0.1 \equiv K.$$

分子進化速度は、定義により、

$$\text{分子進化速度} = \frac{K}{2 \times 3000 \text{ 万年}} \equiv L.$$

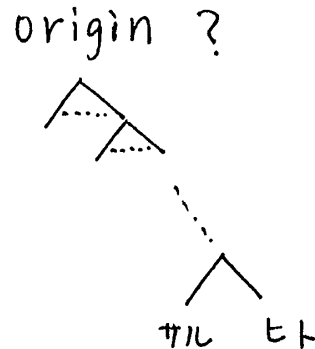
木村中立説の、基本的な考え方は、

‘各タンパク質ごとにLの値は一定’

という原理に基づいている。より厳密には、のちの講議。

中立説から、次のことが原則的に分かる。あるタンパク質について、(1) 様々な種においてその配列を知っている、(2) いくつかの種について、分子進化速度が計算できた、とすると、これらのデータから、様々な種がいつ分岐したか計算できる。

則ち、進化の系統樹が長さ付きの樹木で書ける。



樹木の平坦性問題：今、樹木が非常に大きいとして、無限樹木 T で近似したとする。

問： H をヒルベルト空間とすると、距離的に一様な埋め込み、 $T \hookrightarrow H$ が存在するか？

今、 $T = (V, E)$ を有限グラフ (V は頂点集合、 E はエッジの集合) とする。 $f \in L^2(V)$ に対し、作用素を、

$$\Delta(f)(x) = f(x) - \frac{1}{m(x)} \sum_{y \sim x} f(y)$$

で定め、 Δ をグラフのラプラシアンと呼ぶ。ここで、 $y \sim x$ は、 y は x とエッジで結ばれていることを現していて、 $m(x) = \#\{y \in E : y \sim x\}$ である。

定理：今、無限グラフ T が有限グラフの増大列、 $T_0 \subset T_1 \subset \dots$ で尽くされ、ある $\epsilon > 0$ があって、各 T_i のグラフのラプラシアンの第一固有値が一様に ϵ で下から押さえられている、とする。

この時、距離的に一様な埋め込み、 $T \hookrightarrow H$ が存在しない。

このように局所的に定義された作用素の性質からグラフの大域的な性質を導くことができる。上の条件を満たす T は expandar と呼ばれるものの一つである。最近グロモフにより、次の驚くべき群が構成された。

定理（グロモフ）：ある有限表示群で、expandar を距離的に一様に含むものが存在する。

この群は、これまですべての有限生成群に対して正しいだろうと思われていた、あるトポロジーに関する予想（Baum-Connes 予想）に初めて反例を与えた。

終わりに：この講義では予想されたように、私の答えられない幾つかの分子生物学についての質問がなされた。私に変わって答えて下さった中島裕美子先生（琉球大学遺伝子実験センター）に感謝いたします。

REFERENCES

- [CG] A. CARBONE AND M. GROMOV, *Mathematical slices of molecular biology*, IHES preprint, (2001).
- [ECHLPT] EPSTEIN D.B.A, CANNON J.W, HOLT D.F, LEVY S.V.F PETERSON M.S, AND THURSTON W.P, *Word Processing in Groups*, Jones and Bartlett (1992).
- [G] M.GROMOV *Random wark in random groups*, IHES preprint (2002).
- [K] T.KATO, *Operator dynamics in molecular biology*, IHES preprint (2001).
- [Ki] B.P.KITCHENS, *Symbolic dynamics*, Springer, (1998).
- [宮] 宮田隆, 分子進化学への招待 講談社ブルーバックス (1994).
- [根] 根井正利 分子進化遺伝学 培風館 (五條堀孝、斉藤成也訳) (1990).
- [S] K.STANSFIELD 遺伝学用語辞典 東京化学同人 (西郷薫、佐野弓子訳) (1983).
- [田、村] 田村隆明、村松正實 基礎分子生物学 東京化学同人 (1997).
- [柳] 柳田充弘 *DNA 学のすすめ* 講談社ブルーバックス (1996).